# About Anna and METRO

1:32

Alexey

**This week, we'll talk about product owners in data science. We have a special guest day, Anna. Anna is a domain owner for Data Science at METRO.digital, where she drives adoption of data science and AI topics within the METRO business. And METRO is a large… What do you call it?**

1:51

Anna

Wholesaler.

1:53

Alexey

**Wholesaler. Okay. So, supermarkets, right? [chuckles] Previously Anna led product teams in areas of recommender systems, robotics, and smart logistics. In this conversation, Anna will share her experience working as a product owner. Welcome, Anna.**

2:09

Anna

Hi, Alexey. As you asked me, maybe to elaborate a little bit. First of all, METRO is a wholesaler. Sometimes you think it's very similar to retail – in some aspects, yes, and in some others, no. The difference we have here is just that our customers are B2B. It's not like just you and me come in there and buy something, but in many situations, we expect hotels, restaurants, and caterers to be our customers. But you're talking about data science, so we also expect changes in the behavior of a customer, for example. It's very difficult to inspire them to buy or try out new things. If you imagine a restaurant like a pizzeria – it has these five types of pizza, it will not start buying cherries all of a sudden.

3:01

Alexey

**Cherries on a pizza. That would be interesting.**

3:05

Anna

[chuckles] So that is a really a special challenge for the recommender systems. So, I'm a so-called domain owner, responsible for all data science projects and initiatives at METRO, in the area of data science.

3:20

Alexey

**There were so many times that I tried to get into a METRO store, because it's usually cheaper, but the security wouldn't let me in. Because you're doing B2B, they always wanted proof that I am a business owner, so I always needed to go with somebody who has this proof. [chuckles] Sometimes I actually, when I lived in Poland, they had a MAKRO store, which is also METRO under another name, right? [Anna agrees] So I managed to sneak in sometimes without any documents. Those were adventurous times. [laughs]**

3:54

Anna

That's a great point that you bring up considering that we're talking today to the community. You can think about it like how all people try to get this personalized data – we have it directly. We really know every transaction from the beginning of METRO’s existence is assigned to a customer. So we always know “Anna has been buying those five milks on that day and time.” And as a picture, it’s complete. Many times other companies try to have all the loyalty programs with paybacks and you name it – for sure they're doing it to collect data, but many times there's a certain piece of society who never applies for such programs and you have just a subset. In our case, it's complete and that is always what I try to sell as an advantage for everyone who applied for a job at METRO, like “Look, you have direct data to start working with.”

# Anna’s background

4:49

Alexey

**That's cool. You can’t do data science without data, right? So the main topic for today is product owners and data science. But before discussing that, I wanted to ask you about your career so far. So can you tell us about what you've been doing so far? What has your career journey been like?**

5:09

Anna

Sure, I’m more than happy to share it. I originally studied computer science. Then, in my PhD, I was focusing on – back then it was called data mining, today, we call it data science – and I was focusing on web science. You represent the community as a network, and you try to analyze what makes the networks more healthy or less healthy, and you apply different clustering and classification algorithms to aid that. Then, as I said, you try to say what drives the community to grow or to decline. I was doing this research in this area of open source projects, which is pretty interesting to see – why one open source project evolves and evaluates and others just decline.

5:57

Alexey

**That's pretty interesting. I also happen to run a community, which is pretty relevant to what you're saying. So do you still have some of your research in open access?**

6:12

Anna

I assume my PhD should be accessible online, because it’s the general regulation in Germany. Nevertheless, what I can say about the key point is – you observe this Pareto distribution very strongly. It's not 80/20. In the open source community, it's like 90/10. You really have this very small portion of core developers who are really making a difference and you have many people who just drop a message once, maybe submit a bug report. So it's a very, very long tail and a very dense core.

Back then, for example, I was comparing three projects that were trying to do the same thing. It was BioPerl, BioJava, and Biopython, which are extensions of the corresponding languages – Python, Java, and Perl – in order to do some bioinformatics. You would say, “The application is the same, the community people should be the same.” And one of the findings was, you have to have a strong interlayer. It shouldn't be that steep, like you have three people in the core and just the long tail. You have to have those who are here and they are more active.

If you're just relying on your core – those three to five very central people – once one of them drops out, you have a really big problem. If you have a strong interlayer – something in the middle – there's a high chance that someone from this interlayer will jump in and close this gap. It's not only that. I was able to prove it with data, but I have to say that all these communities were so open. I reached out to them and was asking questions, like, “Look, that is what I see in the data.” They really shared their stories, saying “Yeah. We have observed that. When that person leaves the community, it really goes down with progress and so on and people get demotivated.” So I even got a lot of not only quantitative, but also qualitative, commitments to what I have found.

8:20

Alexey

**Yeah. If anyone who is listening is interested in doing some of this kind of research, and you want to get some data from DataTalks.Club, please reach out. That would be quite interesting to try. I've also heard about a rule called the “90-9-1 rule” where you have a 90, which are people who are silent, 9 who are active, and then 1 is this core. I think this is very similar to what you mentioned.**

8:45

Anna

Yeah, sounds like it to me too. But coming back to my biography – I finished the PhD in this area, but I was always kind of interested in structuring and organizing things. That is why I first went to Zalando, where I was responsible for smart logistics, automation of the warehouses – the warehouse that’s a warehouse, not a data warehouse, but the physical one where you store the items on shelves and then you do the picking/packing and then it's transported. I assume for the people who are listening to us. For example, one of the tasks was the picking process, which is very time consuming and also a workforce-consuming one.

There was a machine constructed to automate the process and I was responsible for the integration of this automated machine into the processes. Because you still have the workforce, which is running around doing some manual picking, plus this automatic machine – so there are two flows, which go apart and you have to bring them together as a packing station, since now you have to bring everything together. At some point in time, Zalando also signed contracts for AWS, so we were doing, like someone said, “Yeah, we need to develop a new piece of software for the warehouse’s automatic machine anyways. Let's go directly for micro services on AWS.”

So I had these two hats – to streamline that process with the task force team to transition AWS microservices, and to integrate this machine. I think that was a very important experience for my life because it was very execution-driven. Literally, if you're working very close to this operation, you can imagine that behind every delay, someone can put some amount of money that you are losing, for example, if the machine is not working. You can put a dollar/euro amount behind every package that is not sent out on time. I learned to work under very high pressure. And don't get me wrong, I’m complaining – I really enjoyed it. But just to give you a number – in the warehouse where I was working at the time, we had up to 2000 workers. Maybe not everyone relied on the machine, but even if you say it's a thousand workers who are not able to work because the system is down, you know how much you pay each of them for every minute. Because we are in Germany, no one will send you home saying “You have nothing to do, we're not paying you.” The people are still there and sitting and just playing around.

There was, really, the management of this warehouse coming to me and saying “Yet another minute. Yet another minute. What does this mean in terms of customer satisfaction and also money that we're losing?” From there, I moved to METRO, where I was a product manager for recommender systems. I guess that very strongly formed what I am and how we managed to really go from this POC situation that you're facing a lot in data science, to really having it as a product. I came from a very strong, driven “Deliver, deploy every day, there's so much dependency.” And then I come to the area of data science, which feels like research. I came and said, “No, no, no. We’re trying to take the best things we know from computer science and somehow put it together with things we need to do in data science.” And I really do believe that we were successful. And I really do believe that we have to use product management approaches for developing successful data science products, which are then in production.

12:49

Alexey

**So you studied computer science – you studied data mining and web science. Do you think it helped you with what you're doing right now?**

12:59

Anna

Yes, I do think so. I still think that in a management position in data and tech, you need to have people with a data tech background. I really do think that it matters a lot for my direction. I can understand the language they speak and I can also understand why, for example, research here and they – that’s still important. It also gives a certain respect, but also, I'm then able to go and advocate for their needs. It's not just for me “blah, blah, I do not understand this,” it’s like, “Okay, I got your points,” and I go back to the stakeholders and say, “Look, people need more time,” or “The data is not sufficient,” or “There is a limitation of infrastructure.”

# The importance of a technical background for data product owners

13:50

Alexey

**Do you think product owners and product managers – we will actually talk about what those terms mean – but I'm wondering whether you think they should have technical background or not? Or is it just a “nice” thing to have?**

14:01

Anna

I guess it depends. I also know some product owners who do not have a technical background, who are really great people. So I guess soft skills are also very important. Also, customer understanding, or customer obsession, can really make sense and make a difference. However, I guess it still depends because there are, for example, very technical products. One example would be – at METRO, out of data science, we have cloud foundation, which is a product about giving you a piece of cloud infrastructure.

I guess you have to be a very technical product manager, otherwise you will not understand what it's all about. You can then also not be that passionate about it. If we are talking more about really customer-facing things – end customer-facing – then you really can connect yourself with the end customers and say “I'm passionate about this product because I am an active user of this app.” So maybe it's not that sufficient.

# What are product owners?

15:11

Alexey

**So what do product owners actually do? Who are they?**

15:18

Anna

First of all, maybe the question is “What is the difference between product owners and product managers? I guess it could also be the definition. [cross-talk]

15:30

Alexey

**What I was saying was “Who are product owners?” It's also interesting to see the comparison – you mentioned that you joined METRO as a product manager, but I think your title is “product owner,” right?**

15:50

Anna

My title today is Domain Owner. Now I'm given the stage app. Back then it was Product Owner. Again, we have readjustments to product managers, and I guess in Zalando, it was the other way around. I see that it’s not that clear of a distinction and maybe someone can write to us in the chat. But I have experienced it a lot in companies where you really don't know what the difference is. Today it’s Product Manager, and then there's some restructuring, and its Product Owner. For me, because I’m sometimes asked by friends, “Okay, coach me. What do I need to do to be a product manager?” So that is my personal definition.

I guess the most important thing is to be able to make decisions. That's what I learned. For sure, the stressful situation was a warehouse, where it's about sending the next 1000 items out or not – it's really a very strong decision you have to make. But even in data science, every time, you need to make a decision. Your data scientists will come and say, “Hmm. The model itself is good. But maybe if you give us two more weeks, we can try to improve it a lot.” For example, being brave enough to say, “I believe it's now good enough. I'm clearly communicating the quality to the stakeholders, but now is the time to go live.” This ability to be brave enough to make decisions is really something that’s very needed to be a good product owner/product manager.

Then there’s being passionate about what you're doing – really loving it – because what I see for me, as a product owner, is someone who sits between stakeholders and the developers (or data scientists). So on one hand, it’s a translator from requirements to the team, but it's also an advocate and a shield for the people. So it's someone who really protects the team in the sense that something like the expectations are not right. How often have I had these discussions like, “How about you take one person and it's ready tomorrow?” [chuckles] That is also the product manager who sits in between and says, “No, we need this and that because the process requires this.” Take a very good example with data science, where people say something like, “Take one FT (one person) and solve the problem.”

18:22

Alexey

**What does FT stand for?**

18:25

Anna

Something like a resource. You can Google, I don't know. [chuckles] Then I am the one who goes and negotiates and says, “You need someone who will orchestrate the whole machine learning flow. You need one data scientist. You need some DevOps or maybe a data engineer.” And it's not right that you just say, “Here, you take it from there and you, data scientists, do it now.” To summarize – being able to make decisions and risk, to protect your team, to be a good interpreter from stakeholders to your team, and then from team back to the stakeholders, and last but not least, being really passionate about what you're doing.

19:06

Alexey

**What you described right now is a product owner, product manager, or both?**

19:13

Anna

For me, it's the product owner. I guess people put more execution into the product manager role. A product owner is more like the CEO of their product. So, I tend more to be on the product owner side.

19:28

Alexey

**By “execution” you mean… [cross-talk]**

19:30

Anna

Because a product manager I guess, is the one who does more managing – streamlines the process, takes less ownership. So that is where I see the difference.

19:42

Alexey

**So this is what you mean by “execution,” right? Meaning they manage the process, making sure the team is working fine – all these rituals like scrum sprints, estimating – all these things are usually done by product managers?**

20:00

Anna

You still have agile coaches or scrum coaches or whatever they’re called today to follow up with the rituals. But I guess in a nutshell, and as I said, there's not always much of a difference. It depends on who you ask. But in my eyes, product owners are like the CEO of a product, while product managers maybe have less power, and maybe have to listen more to the tech lead and the teams. But I'm more than happy to hear what the audience thinks.

# Product owners vs product managers

20:34

Alexey

**I'm just curious. I did work with product owners. In some companies, they have only product owners, while some companies have only product managers. In the company where I work right now, OLX, we have only product managers – we don't have product owners. I'm just wondering, in this case – if there is no product owner officially in the team – then the product manager wears this hat, right? [Anna agrees] So they make these decisions, like you said, “Okay, we need to ship it now,” or “We need to wait for two more weeks.” Also, you mentioned that product managers work together with tech leads, right?**

21:16

Anna

Yeah. Correct.

21:17

Alexey

**So they both can be owners at the end.**

21:20

Anna

Owners, correct. And in your company, Alexey, what is then the next in line? To whom do the product managers report?

21:29

Alexey

**Head of Product.**

21:30

Anna

Head of product. Okay, in our case, it’s like Domain Owners. But again, that's just the naming.

21:37

Alexey

**In the chat, I see that FTE stands for Full Time Equivalent, or Full Time Employee.**

21:45

Anna

Yeah. Full Time Employee. Yeah. Normally, I have these discussions a lot, like, “Can we have the budget for one person, but this one person can cover three use cases from data science?” Then it's part of the task also of the product owner to say “No, it's not doable.” And explain why.

# Anna’s work on recommender systems at METRO

22:08

Alexey

**I think you described, more or less, what you did. You said that you were taking care of recommender systems at METRO digital. So what did you do there as a product owner? How did you exactly “take care of recommender systems”?**

22:25

Anna

When I joined METRO, there were like one and a half people [chuckles] more or less. It was more a kind of research environment. I mean, we were getting requests here and there from a dedicated country, which now plans to have a campaign in the next month and they say “Okay, there’s also products that we would love to put on a special offer. To whom should we send out a newsletter?” And that was pretty manual, because it was like – you get this list of products, which maybe were somewhere in the database, but it was like manual. “Here's a product. Here's the target groups we would love to face.” There was collaborative filtering in between but it was really – there were requests, manual processing, and the results generated and, again, shared, written somewhere in the database. When I came up, I said “Okay. This doesn't scale.”

With this, first of all, how many people should we hire? Back then we were operating in 25 countries, and meanwhile, we are in 23. But otherwise it's like you have this highly paid data scientist spending half of their time actually exchanging emails back and forth, “Which group should be picked?” And just clicking on a button waiting until this was calculated. And I went and said “No, no. I want to have a use case where we clearly serve an API and we do real software development here (as I was used to from Zalando).” So firstly, we introduced the whole agile setup – with stand ups, with two week sprints (where we tried to really set a goal for two weeks). In terms of processes, we changed it a lot, but also in terms of our target group.

To be honest, we still support the use case of the newsletter, but we automated it a lot and dedicated the support of two people who do nothing but this execution. I said clearly, “Okay, where are the use cases where we can use the endpoint?” And that is for sure our online shop. So as you go to the online shop, you're looking at a banana, for example, and underneath, there’s a “frequently bought together” field – so people who were buying bananas are also interested in this and that. So I said, “That's great. We just provide them an endpoint, and they integrate in it.” Back then I guess (METRO online shop) was present in like 10 countries. Then they went up to 12. Then they went up to 18. For us, because we trained our model already for all 25 countries, every country that went live – we were already live there. How great is that? So that was really scaling.

Then there was just one flow created (one ML flow), there was an API, we were able to set Datadog on it, to see how reliable it is. We also didn't take care of anything – our endpoint just says, “Okay, given a product and the customers who are looking at it, what are the best six frequently bought together products?” So we just submit the results in the form of IDs and that the rest of this is the responsibility of the customer-facing solution to grab the right pictures, descriptions, and the right price. That is also great, because then, you as a data science team, really focus on the core task. You're not trying to see “Okay, do I have a description (or whatever)? Can I consider the results complete?” Because many times if you just do the manual process, you're consuming party stats, but you’re also thinking “Yeah, can you also search for the pictures and submit the results with the pictures?” Or whatever. So this example is one where we also took a clear step.

Then we extended our portfolio of events to alternative recommendations. That one is – if you're looking at a product and it's not available (because we don't want our customers to be disappointed) Alexey you said in the beginning that you have been to METRO so you know it has a huge assortment. Normally, if you don't have *this* dedicated milk, there are at least 50 other milk types that will have the same percentage of fat, the same size, and everything else. So it really makes sense just to propose a reasonable alternative to do it in a personalized manner. So that is another one. Once something’s out of stock, it says “We have the following alternatives for you.”

And last but not least was the recommender. Again, when we started, I told you “Okay, give the special offers to whom it's relevant.” And the whole time I was thinking, “But I get more out of it. I get more out of it. Newsletters are a manual process, but where can make it more automatic?” And it's our app – we have M Companion, which is an app of METRO. You go there and there’s several tabs and one tab is a special offer. It was just that all products that are under offer were listed there. Let’s say, for example, you have a vegan restaurant that doesn’t sell any meat, but because we have meat as a special offer, you're getting meat on the top. I guess many customers got kind of pissed, like “Why are you offering it to me? It's not relevant.” Here, we applied it like – every time the offers are updated, we take the offers and we train our model to whom 200 top recommenders rank in the order of relevance per customer. When you go to this app, and click on “offers,” your list will be different from mine.

And again, M Companion rolls out to a new country, here we are – it's an endpoint, it's directly solved, we can update it, we are independent, we can run an A/B test without creating any problems. That is the approach that I would recommend everyone to do.

29:05

Alexey

**So you mentioned three use cases. First is a newsletter – so all the clients of METRO receive newsletters with personalized recommendations. I think you said that you segment first, and then each segment gets the same newsletter, more or less, right?**

29:23

Anna

No, no, no. It's really personalized. It's a really customer-focused product. You can, so why not?

29:29

Alexey

**Okay. Then you have the online shop, where you have a recommender system. And then finally, in the app. So three places.**

29:37

Anna

Yeah. But I would actually count it as because “alternative recommender” and “frequently bought together” are really two different recommenders. One is cross sell – I try to encourage you to buy more, so it’s really saying like “You're looking at a banana, also buy an apple.” Whereas the alternative is just, “You're looking at bananas. These bananas are not available, take another banana.”

# Expanding the data team

30:01

Alexey

**Okay. I'm just curious. You said that when you joined METRO, there were “one and a half people” and then you said that you need software engineers, you need agile processes, etc. So who was on the team and who did you need to hire?**

30:20

Anna

Yeah. It's a very great learning experience. I always love to share it. So I started first on the role list. Our very first position description, we were looking for people with PhDs in mathematics and statistics. I still do believe you need those, don't get me wrong. But the question was about the amount. It depends on how far you are in the process – sometimes you're still in POC and trying things out, but once you're pretty ready with the model and you are just improving it, it's enough to have one data scientist per use case, maybe.

We pretty quickly recognized that if you really want to do this productionized software, something which runs into production, as I said, 24/7 – where we have the whole infrastructure, the Datadog, the ML flow, and so on –we need real engineers (today they’re called machine learning engineers and sometimes MLOps – depends a little) I would say from today's perspective, you need both roles and it's a little different, because ML flow is really like “How do you structure the flow? Some tail, bring it there, retrain, risk, and so on,” while MLOps is being able to create an ideal infrastructure for machine learning projects. Sometimes I get asked, “Okay, what is the ratio?” Again, depending on how far you are in the prod process. At the beginning, you can start with one to one.

32:00

Alexey

**So that’s one ML engineer one and one MLOps and one data scientist?**

32:03

Anna

Yeah, and one data scientist. That should be your basis. But once you're going more and more into production – and, as I said, there may be just some improvements of your model, but it's also all about being stable and running – then it can be up to three engineers for one data scientist.

32:24

Alexey

**Okay. Let's say you don't have money for three people – who will you hire first?**

32:31

Anna

[chuckles] Again, it depends where I'm staying. If it's really a completely new use case, it may be data scientists. Ideally, I would go for a generalist. So, again, if someone listening now to us [cross-talk]

32:45

Alexey

**Full-stack, right?**

32:46

Anna

Yeah, if you really want to be the kind of the person everyone wants to have, try to evolve yourself to the generalist role – someone who can at least set up some basic infrastructure, create some ML flow *plus* being interested and understanding data science. That is where I would start. I can also be very honest with our audience, because I guess it's not a secret – for some initiatives, when I don't know if I should hire someone-long term, I would also reach out to some external support. With external support, you can reduce and increase pretty dynamically. Once I know “Okay, that is something I would love to go for.” The business really shows. It's not like I say it. Maybe let me take it back. If you really see that it improves the business and there is a potential, then we go to hire people.

There's another approach we also use and it may also be a recommendation towards our audience. I found that many people – today I have already 22 directs – are always willing to learn something new. You can also ask around in just the existing resources, if they would love to have something like 20% commitment, or maybe even 50% commitment, for the next three months to run a POC. It’s a great thing to give back to your people, just to say “Look, that's another use case.” That is also something they can put on their CV. They can say “I was creating a model for recommenders and markdown.” That's what we have done, so maybe someone who participated in both is listening now. And I got really great feedback from people, like “Oh! I learned a completely new piece of data, a business partner, (and so on, so on.)” Then after you have done it in this setup, you can then say “Okay, those are the people I need,” and you decide whether you hire or you relocate.

# Types of algorithms used for recommender systems

34:53

Alexey

**There is a question from Valeria, “What kind of recommender systems (or probably algorithms) did you use? Was it collaborative filtering or was it different for each of the use cases you had?”**

35:03

Anna

Yeah. It's different for each of the use cases. Collaborative filtering we were using for scoring already. Let me be clear here – it’s been one and a half years that I've been in this domain owner position and not working that case. Maybe my information won't be up to date. But we also used a modification of Word2Vec to find combinations from people – to become able to score the relevance of products to people.

35:38

Alexey

**This Word2Vec is for things that are frequently bought together? Like apples and bananas?**

35:45

Anna

Yes, correct. There is a special modification. Normally you would expect that it tries to use semantics, but that's also like learning semantics out of your transactional data.

# What kind of knowledge and skills data product owners need to have

35:55

Alexey

**Did you actually need to get into technical details for all that as a product owner? Did you need to know that, “Okay, this is *this* specific Word2Vec modification?” Or was it something that you just wanted to learn because of your background?**

36:11

Anna

You should understand the basics and you should be able to challenge it with the right questions. For example, what we have done and kind of experimented at a certain point of time. You know how people who have done a PhD in mathematics, or statistics, or computer science – they got used to reading some scientific papers. And there are great scientific papers outside. Here and there was always a situation of someone coming and saying “Over the weekend, I was reading this paper. I think it's a great idea. Let's try it out.” I really wanted to show that I value this ambition, and this time investment. But I also didn't want us to jump on every new paper, like, “Let's implement it!” So one of the learnings was, at some point, we created just a basic table in Confluence, where we were collecting all these ideas. So you place a link to the paper and you write a short summary. But the important thing was that there were two columns – one column asking “What is it good for? Is it good for making a good collation quicker? Or is it good for improving the quality (improves accuracy)?”

Coming back to your question, it is very important that you understand that you can ask these questions and challenge people. It sometimes gives really good feedback to the people – they recognize, “Oh, you're right. It’ll only speed up the process.” And you say, “Okay, now we train a model once a week. Do we even need to speed it up? Will it make a difference? Today it runs for two hours, and then it will run for one hour and a half. Will it influence the business?” And then people are like, “Okay, you're right. It will not.” So I don't think that I have to understand (or any product manager has to understand) in very great detail. But you have to understand data science principles and what it’s about, what it means. What does a metric mean? What does quality mean? What does any other type of improvement mean?

38:32

Alexey

**Now you work as the data science domain owner. How is it different from a product owner? I think you answered that question, right? You are now a manager of product owners, right?**

38:45

Anna

Yeah… no, no, no, wait. There’s a difference. It's not like a Head of Product – it's a Head of Data Science people. So the people who have the role of Data Scientists or Data Analysts do report to me, but they are then part of a product team. And product people do report to the Head of Product.

39:09

Alexey

**Okay, so it's like the Head of Data Science in our case.**

39:13

Anna

Yes, Head of Data Science within an engineering organization. The idea is that I am kind of sitting there in the middle, so I can ensure that we are not doing the same thing five times because it's somehow streamlined over my head. I also play an umbrella. I bring the data scientists from this team together with data scientists from this team and say “Look, it seems that you're doing similar stuff, maybe you should exchange ideas, or maybe just support each other. And I can enable what I just described – this rotation –to say “Alexey, we have a new initiative in markdown. Do you want to try it out?”

39:57

Alexey

**Interesting. But you said you manage 25 people?**

40:01

Anna

22

40:02

Alexey

**22. Are they your direct reports?**

40:05

Anna

Yes.

40:06

Alexey

**So you actually need to do performance review, all these appraisals, everything like that – in addition to knowing all these domains?**

40:16

Anna

Yeah. It's a lot. A certain amount of understanding is still required – people do expect me to be able to answer things like “You’re the data science expert. Tell us.” We have eight big data science cases, and they're really a broad range. Even though I have a great focus on recommenders, where I can also go in great detail because it's where I originally came from, but as I said, we have a markdown initiative, like “What is the optimal discount given the expiration date of this product?”

We also even give credits. I don't know if you know about it, but at METRO, as a big customer, you can get a loan like you would get in a bank. For this we do credit scoring. It's completely new – another area. We’re also trying to do some dynamic pricing, and we do assortment optimization. It's challenging to be up to date everywhere, but as I said, I try. I try hard plus I assume I know what questions to ask.

# METRO’s markdown system

41:34

Alexey

**What does markdown mean? Because in my world (for me) markdown is what I use in GitHub when I create a document. This is a markup format. But I don't think you’re talking about that. [Anna agrees] So what is “markdown”?**

41:49

Anna

It's actually a great use case. You know how the whole world today is talking about sustainability and waste reduction? It goes in this direction. So what your model is doing – you know how many products are left with a certain BBD (best before date). For example, you have 100 of your goods left of this type, and they’re expiring in 3 days. What the system then tries to calculate is “What is the optimal discount we should give, so that we will be able to sell out all 100 of the goods without losing too much money?”

In terms of what is happening, I can even explain the concept of a model. It’s more or less based on elasticity – price versus how much people are willing to buy it. You say “How much I predict to sell at a price of certain level, and you go through all the created prices, and you take the minimum discount, which brings you to a series that sells everything out.” And there's three targets underneath – you certainly want to reduce waste. In some countries, we even put the food as a donation, but we still pay money on top. So that's really a penalty that you have to pay. So we reduce the waste. Why just throw things away?

The second thing is to increase our revenue. If we maybe start just three days in advance when doing the discount, and you manage to sell everything – you may have almost lost nothing. Last but not least, there’s also streamlining of the process. That is also very interesting, and I would love to share this with the community. Sometimes the automation that data science brings, also brings a clear process. The feedback that we got about markdown in the beginning was not about the numbers. It was about the automatic process. Now it's a system.

There's an app that’s called “M Store” (METRO store) and the employee runs with it and says “Okay, now you have to discount this yogurt by 20%,” and then the employees are just doing it. Before it was a process where an employee had to follow some guidelines – the guidelines were somewhere, so every time you had a new employee, you had to onboard him or her. Also, depending on the store there was also fraud. Fraud in the sense that “Hey Alexey. You're my friend. You just come and I put 50% stickers on this yogurt and you take them.” That is also something that was taken out. It's just that there's now a clear process and clear guidelines – the person doesn't even have to think about it.

# Problems and ideas should come from the business

44:48

Alexey

**Was it you or your teams who needed to think about this process? Or did you have help from the actual operations people – people who run in the stores. What did it look like?**

45:03

Anna

It was the business that came with the proposal. In my eyes, it's great if you have a sole visionary and you have great ideas, but I truly believe that business operations are your customer who knows where the problems are or what the problems are that they want to solve. And that is where we should start. I always say, depending on whether it’s data analytics or data science – I go to the business and say “Tell us what your problem is. Tell us what you would love to have improved.” And that's where you have to start solving. But sometimes there also can be challenges.

# How Anna handles all her responsibilities

45:57

Alexey

**I lost my train of thought. I wanted to ask you, “What does it actually mean to be responsible for a domain?” You have six domains, I think, according to your LinkedIn. [Anna agrees] These domains, I think, you've talked about all of them (or most of them) – recommendation systems, churn, markdown, fulfillment, royalty pricing, and you also mentioned this new initiative, which is credit scoring. [Anna agrees] Then it's just one person who needs to oversee all of that, which is you. So what does it mean to be responsible for a domain? How do you even manage to look after so many things?**

46:48

Anna

The data scientists or the data products are still part of the product team. For example, in the example of markdown, there are three teams responsible for the M Store application. The employees are using it to receive new goods, report some damages, and so on. And that's just normal, three product teams with developers who are developing frontend, backend and so on. While markdowns is just yet another service. So that means that at the moment, I have two people and some external support that are responsible for this piece. But they have a product manager who streamlines that and pushes it forward locally. So it's not that I have to oversee it on a daily basis.

I still need to understand what’s happening there because it goes hand in hand with people management. So if my directs give some feedback or complain, or raise some proposals, they sometimes want me to get pulled in. Sometimes I’m also happy to challenge something based on my experience. For example, if they say, “We would love now to improve this and that,” I will be pulled in. Also if there is a proposal to try something completely new, again, it's me being there.

I don't need to oversee the process of the dailies, retros, and so on – those happenings have dedicated product teams. It’s the same for recommenders – they have their own product owner. It’s the same for the churn and CRM – they have a product and Agile coach. They work, but I'm the one who can rerotate people, who oversees it in a rough detail, who is called if there needs to be a justification.

48:41

Alexey

**Do you also manage budgets?**

48:43

Anna

Yes.

# The process for starting work on new domains

48:44

Alexey

**Okay. Because you mentioned FTE, so you probably operate in these terms, like “Okay, for this project, we have two and a half FTE. For this project, just one and a half.” [Anna agrees] That's what you do, okay. When I was preparing questions, it was just six domains. Now, you also mentioned this new one. How does it happen that you get an entirely new domain? What's the process?**

49:14

Anna

So you mean an entirely new use case? As I said, in many ways. Sometimes it's that the business will come and say “We have a great idea!” And that's really where I have to be active. First of all, I have to justify “Is the idea really big enough?” Because, I will be honest, I have seen situations where it was like “Oh! AI! We have some text issues in our data! Can we use AI to correct it?” And so I was like, “Okay. What are we talking about? Some broken entries? How many broken entries are we talking about? 200? You're not doing any text AI for 200 entries. Hire yourself a student worker for two hours, and let him or her correct manually and you're done.” This is the kinds of conversations that I'm driving. That can be one that I push back to the business.

Nevertheless, there are some great ideas. For example, with assortment there was a situation. The business side came and said, “We have great ideas. We need to reduce the assortment. That is part of our big journey of a METRO and we would love to optimize assortment (maybe not reduce, but optimize) to see what is driving our customer into the store.” In a data sense, it's a great task, and I was very, very inspired. So I said, “It's great. But then bring me some money for people.” Sometimes it also stops there. You say “bring the money for people,” and then I hire you the best talent, I ensure we have the right setup, but we need to hire new people. That doesn't sound like a “20% on top” kind of project. Then for example, if that happens, then you can start running it. As I said, in a nutshell, I always prefer if the idea comes from business.

51:23

Alexey

**Interesting. I really love this – that you can just hire somebody to do this, because it's just 200 rows, right? Somebody can just go through this. I think this is what I, as a data scientist, have problems with. For me, machine learning is a hammer and I want to use it for every nail. But then sometimes manual work is – it's boring, but it gets the job done. That's interesting.**

51:58

Anna

Sometimes, maybe. Let’s add something here. Sometimes I would agree to do some manual effort, but manual not in the sense that – correcting 200 lines, it will be 200 entries, it will never get a go. But sometimes people will come and say, “I would love for you to optimize the assortment. I'm not sure that if you're able to do it (like you as a data scientist, not me personally),” then I'm okay to say, “Do something more at the level of Jupyter Notebook and then the results are just here in CSV. Then let the corresponding category manager check the results.” It's completely not the way I want to have it. Normally I'm [audio cuts out] CSV, Excel, whatever. But to get a foot in the door, I would sometimes agree to do this kind of small proof of our abilities that will get this buy-in. Really, with clear communication, it is not the process. We just can share with you the results and use and check.

# Product portfolio management

53:09

Alexey

**Okay. I see a question from Jamie. By the way, do you know what product portfolio management principles are?**

53:19

Anna

We don't have anything like that. I know that some companies like SAP, where I have a friend who is a portfolio manager there. But we don't have things like that.

53:30

Alexey

**Because the question is whether you use this for determining the products that you should have, and how much resources to put there – if I understand the question correctly.**

53:43

Anna

Yeah, I guess with the portfolio I mean, you would say that now there are eight use cases I named. So that’s my portfolio and you somehow decide. I have created something like that – it’s just a list of questions that I always share when someone approaches me with an idea. Obviously I have some structure there. But as I said, from there on, it's more of an overtime development process. It’s like “Take this commitment from some internals with this 20% overtime, see if you maybe bring in some externals and then see how it goes.”

# ProductTank and Anna’s role in it

54:21

Alexey

**I see that we don't have a lot of time, but there is a lot more that I actually wanted to ask you. One thing I'm really curious about – I know that you're one of the co-organizers of ProductTank in Dusseldorf. I’m wondering what this is. What is this thing called ProductTank?**

54:42

Anna

ProductTank. There's a very big community called Mind the Product. It's originally from London. The initiative of the whole Mind the Product Community is to grow product management. You can get a co organizer of ProductsTank in your city, so you officially belong to this cooperation or you’re franchising under their name. And that is a meetup you try to run regularly about product topics. We have, for example, one on Thursday the 27th of October. We try to bring like product folks, but also everyone else, talk about how to create great products. For example, next time we'll be talking about hypothesis-driven development via testing and automation.

55:44

Alexey

**That's an offline meetup, right?**

55:48

Anna

It is offline. The next one will be offline again. For two years we were trying to do it remotely. I guess, the first two remote ones were real successes. We even managed to get some product owners. There was one lady connecting from New York and giving a talk and we were so motivated, like “Oh, now you can have speakers from all over the world!” But I have to admit, after the two first sessions, I had the feeling everyone was tired. Now again, you're already in your meetings the whole day online for eight hours. And then you have to sit down and listen again to yet another speaker. So we were not doing much last year. But this year, it’s again a live event.

56:41

Alexey

**So what we're doing now is contributing to this “After eight hours of meetings, now you sit down and listen to us.” But maybe it's not a bad thing in the end. If it's evening for you and the end of your working day, it’s not a bad way to finish your Friday.**

57:00

Anna

Yeah, that’s for sure. But I guess the difference is that here, the people here are willing to watch and educate themselves and see another opinion. A big part of the product community was the community – this new network and just being together, eating this food together. That is a piece that kind of got lost. That is also, I guess, what people are looking forward to.

57:27

Alexey

**Did you ever ask yourself why it's called a tank? Why ProductTank? Why specifically “tank”? Did you wonder?**

57:38

Anna

No. [chuckles] But now you're asking. I have to check.

57:43

Alexey

**Are you checking right now? [laughs]**

57:45

Anna

No, no, no. I have to check at some point.

# Anna’s resource recommendations

57:48

Alexey

**Okay. Yeah, please share what you find. I’m really curious why they thought this was a good idea. I mean, probably tank is not only like a war tank – it probably has multiple meanings. Anyways, there’s another thing. There was a suggestion from one of our listeners, Johanna, who suggested that we should ask every guest for some resource recommendations. And I was wondering if there is any book or other resource that you can recommend to our listeners.**

58:24

Anna

Considering we started with product management and data science, I really recommend one book that’s called Data Science for Business. If you're already a very dense data scientist, it's not the book you should read. But if you're more of an experienced product manager who would be willing to be a product manager or product owner of a data product, you should go for it. It's called Data Science for Business and it's by Foster Provost and Tom Fawcett. If you Google “Data Science for Business,” it’s the first thing you get. And it also says “what you need to know about data mining and data analytic thinking”. It really gives you a very good overview of the topic and what it is about. It starts very basic. It explains what predictions are. And it's very entertaining – very easy to read.

59:38

Alexey

**I think this is not the first time I have heard about this book. But you said it won’t be very interesting for people who already know data science.**

59:49

Anna

It’s more like business people transitioning towards data science, not the other way around. I don't know. Maybe someone who is in data science and wants to understand how business thinks? It will also help you to see, “Okay, that is a perspective of how to explain to them. So that is maybe the way I should talk to business people.” That might be also helpful. What I like about it is that it’s easily written. You're not falling asleep after two pages.

60:25

Alexey

**[chuckles] Okay, that's a good quality for a book, right? Not a textbook. Okay. That's all we have time for. So thanks a lot, Anna, for coming, joining us, sharing your experience with us. Thanks, everyone. There is a very lively discussion in live chat. So thanks for joining and engaging in this discussion. That was fun. Thanks a lot.**

60:49

Anna

Thank you for having me and have a nice weekend.